DOCUMENT RESUME

ED 211 590 TM 820 018

AUTHOR Cowden, Peter
TITLE Private Influences in the Public Governance of Education: The Case of Private Testing Firms in American Schools. Final Report.
INSTITUTION Huron Inst., Cambridge, Mass.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
PUB DATE 81
GRANT NIE-G-79-9-0186
NOTE 63p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Accountability; Achievement Tests; *Government Role; *School Districts; Standardized Tests; *Test Selection
IDENTIFIERS Consumers; *Testing Industry; *Test Publishers

ABSTRACT

There are no formal political means of control or accountability for private testing firms to local school districts. The role of school districts as consumers of these firms was studied. A description of the method of this study was divided into three sections. These included site selection, conduct of field research, and recording and analysis of data. The kinds of test school districts bought would influence what private testing firms produced. The standards by which districts selected and purchased standardized norm-referenced educational achievement tests were examined. The marketplace, rather than formal governance arrangements for meeting local school district testing needs, was discussed. Formal government agencies at the national level were limited in the extent to which they could claim formal responsibility for developing a national test of educational achievement. Since formal authority over education is the responsibility of the state or local government, its national role would be constrained in the development or regulation of a national test. It was concluded that in the case of achievement testing, the marketplace may provide a degree of public control that formal public decision making and politics cannot. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

The Huron Institute

ED211590

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

> This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 820 018

THE HURON INSTITUTE 123 MOUNT AUBURN STREET, CAMBRIDGE, MASSACHUSETTS 02138

FINAL REPORT

PRIVATE INFLUENCES IN THE PUBLIC GOVERNANCE OF EDUCATION: THE CASE OF PRIVATE TESTING FIRMS IN AMERICAN SCHOOLS*

Peter Cowden

The Huron Institute

Cambridge, Massachusetts

Fall, 1981

---

* This study was supported by a grant from N.I.E. (Grant No. 79-9-0186). Thanks are due to Lana Muraskin for her patience and help on this study. Jack DeSanctis and Walt Haney also made many helpful comments on an earlier draft. Special thanks are also due to David Cohen for his suggestions and encouragement on several earlier versions of this paper.

TABLE OF CONTENTS

INTRODUCTION 1

HOW TESTS ARE BOUGHT 6

Why and When a Test is Selected 7
Who Is Involved in the Selection Process 9

THE BASIS OF SELECTION 14

Selection Considerations Unrelated to Norming Methods . 21
Test Attributes 22
Services by Test Publishers 27
State and Federal Testing Requirements 30

THE INFLUENCE OF SCHOOL DISTRICTS AS CONSUMERS 33

THE MARKETPLACE AS A MEANS OF LOCAL CONTROL 36

THE MARKETPLACE VS FORMAL GOVERNANCE ARRANGEMENTS 38

SUMMARY AND CONCLUSIONS 43

METHODOLOGY 47

REFERENCES 55

INTRODUCTION

Educational testing has increasingly come under public and professional scrutiny. One only has to look at emerging social developments such as "truth in testing" legislation, recently passed in New York and now being considered by other state legislatures and the U.S. Congress. A National Consortium on Testing comprising more than three dozen national educational organizations has also recently been established to review standardized testing in detail. The American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME) have announced another joint committee to revise their "technical standards on testing"; this represents their third revision in the last two decades (Haney, 1978). Such developments have led one test developer to conclude that "the going is getting tougher for educational testers" (Boston Globe, 1978).

In part, increased public and professional awareness can be explained by the widespread use of these tests in the last two decades. The sale of standard tests had grown fivefold since the early sixties (Association of American Publishers, 1980), and now amounts to an annual sum of over 200 million dollars (Kohn, 1975). This boom is due mainly to the growing use of tests by local and state educational agencies. At the local level, there is evidence that tests are commonly used to gauge the success of individual schools and the performance of teachers (National School Boards Association, 1977), as well as to identify and manage individual student differences in the school and classroom (Levine, 1976).

Other studies report an enormous increase in state-sponsored testing, both for determining school districts' advancement (Kemble, 1976) and for setting minimum competency standards for graduation (Haney & Madaus, 1978). One study in California reported that 55 percent of the schools made curriculum changes as a result of a recent statewide testing program. Others have pointed to the increase in testing at the state and local levels to meet federally mandated or federally supported assessment activities (Clasby, 1973). It has even been suggested that standardized tests are now so thoroughly ingrained in American schools that "it is a sign of backwardness not to have test scores in the school records of children" (McClelland, 1973).

One mainstream of concern about educational testing focuses on test makers--private testing firms. This concern stems from a recognition that private testing firms, unlike other private organizations in education, are unregulated through formal governance arrangements. That is, they are held accountable neither to those elected nor to their representatives (Cohen, 1979). There is some truth in this view. Most other private and quasi-private organizations--for example, professional associations (like the American Federation of teachers or the National Association of Elementary School Principals) and public interest groups (like the Children's Defense Fund and the National Association for the Advancement of Colored People)--are held accountable in their lobbying efforts by elected officials or their representatives at the federal, state, and local levels. This is evident in accounts of such formal public decision-making processes as the U.S. Congressional legislation of categorical

programs (Bailey,1965; Thomas, 1975), state legislative setting of school curriculum (Marconnit, 1968; Bowles, 1966), and local school board hiring policies (Gilbert, 1966). These organizations are also held accountable by judges (appointed representatives of public officials) in their court decisions on matters such as school finance reform, education of the handicapped and desegregation. While views differ about the extent to which these processes act to hold private and quasi-private organizations accountable to the public,* private testing firms are not accountable through any formal means to elected officials or their appointees--though this does not necessarily mean that they are not accountable to the public.

One school of political-economic theory suggests that the public as consumers can hold private firms accountable through the workings of the marketplace (Friedman, 1962; Dahl, 1953). By deciding what tests to buy--and not to buy-consumers can be seen as exercising their political preferences, just as others may exercise theirs by voting in elections. Like elected public officials, private enterprise competes to meet public needs and demands. If it fails to meet those needs and demands, consumers will not buy its products and it will suffer from a lack of public support. According to this view, then, school districts acting as consumers can

---

* For example, one school of political theory argues that private and quasi-private organizations are self-appointed public representatives who have the advantage of competing with more diffuse constituencies of elected officials and thereby can exercise a disproportionate influence over public decision making (see Madison, 1961; Bendix & Lipset, 1967; McConnel, 1967; Kariel, 1961). Conversely, pluralistic political theory argues that these private influences are part of the democratic process, as evidenced by the conflicts and compromises preceding consensus in public decision making (see Dahl, 1953; Truman, 1951; Lindblom, 1977).

hold private testing firms accountable through the marketplace.

Nevertheless, concern about the accountability of private testing firms persists. Some argue that because these firms are monopolistic and lack competition, they show little daring in the product they offer their consumers (Kohn, 1977). In this view, they are portrayed as public utilities but without public control. A related concern is that private testing firms offer their consumers a uniform product that provides little opportunity for real or significant choice (Hoffman, 1962; Levine, 1976)--perhaps as one might see major car manufacturers such as General Motors, Ford, and Chrysler, but without international competition. Others suggest that private testing firms have enough influence with government officials that few efforts are made to regulate or monitor their activities (Karier, 1972, Kiersh, 1979). This view resembles charges made against the food industry despite the existence of a Federal Food and Drug Administration (Hunter, 1972). Others still feel that private testing firms' interest in profit has eclipsed their responsibility for ensuring that their customers know how to use their products correctly (Oles, 1977)--perhaps like a toy manufacturer who sells a gun that can harm children if it is improperly loaded. A further charge is that private testing firms sell their product under false pretenses--claiming that their test can provide information that it cannot. For example, there is some question whether educational achievement tests do in fact measure achievement or are simply another measure of ability (Levine, 1976).

While the list of concerns about private testing firms' lack of public accountability is long, there has been little formal study of the

issues they raise. An important issue that cuts across many of these concerns is the extent to which private testing firms are in fact held accountable through the workings of the marketplace. Do consumers influence and control the tests produced by these firms, or do the latter exercise sovereign control over the tests they produce and sell? What attention has been paid to this question has been focused on admissions testing for colleges and professional schools and on the main provider of such tests, Educational Testing Service (ETS)(e.g., Nairn, 1980; National Education Association, 1977; Strenio, 1980). Considerably less attention has been given to this issue at the elementary and secondary school level; yet most of the nation's 17,000 (circa) school districts are reported to buy standardized norm-referenced educational achievement tests from private testing firms (National School Board Association "standardized Testing" Research Report, 1977). The extent to which school districts as test consumers influence private testing firms is particularly important to examine, because elected school officials or those reporting to them are invested--through statute or tradition--with authority over education. Should they lack influence as consumers, this might indicate a weakening of local control and democratic responsibility and an undue and unaccountable exercise of power by private testing firms.

This essay focuses on the role of school districts in the marketplace. It will assume that school districts are rational consumers--that they will buy the test that meets their needs and wants, much as drivers buy parts for their car--with a specific need and purpose in mind. Thus, through the kinds of tests they buy, they will influence what private testing firms produce. But this assumption will also be examined, since to the extent that consumers do not know what they want and need, their influence

in the marketplace may be limited. For example, school districts might buy tests as consumers buy girl scout cookies--with a sense of fulfilling a social and cultural obligation, and with little attention to the product as such. In this case, their influence on private testing firms is likely to be limited. To explore the role of local school districts as consumers, then, this essay will examine how and on what basis local school districts select and buy standardized norm-referenced educational achievement tests. Answering that question will have some important implications for assessing the effectiveness of the marketplace as a means of local school district control and influence over the tests they buy. The essay will conclude with a discussion of the marketplace versus formal governance arrangements for meeting local school district testing needs.

HOW TESTS ARE BOUGHT

How local school districts buy tests is an important question to answer because it illuminates their behavior as consumers. In particular, it provides a way to assess whether they buy what private testing firms persuade or tell them to buy. Lindblom (1977) calls this "circularity" in the marketplace, because the needs and wants of consumers are determined by those selling the product. A classic example of circularity in the marketplace is that of the door-to-door salesman selling a housewife a brush she does not need: her decision to select and buy a brush is based less on her own need than that of the seller. Similarly, if local school districts tend to buy what they are told to buy, they can be assumed to be less than rational consumers and to have little influence in the marketplace--just as could be expected in the case of the housewife.

In order to assess the extent to which private testing firms may enjoy this position in the marketplace, several preliminary questions about the behavior of school districts as consumers need to be answered. These include why a school district selects a test in the first place, and what the relationship is between school districts and private testing firms in the selection process. The following sections answer these questions by describing why and when a test is selected and who is involved in the selection process.

Why and When a Test is Selected

Teachers, principals, administrators, and school board members commonly agree that their school district enters the market for standardized norm-referenced tests mainly for one reason: to inform their community on "how their school district and schools are doing compared to the larger population" (Interview with Director of Public Services, 12/13/79), p. 1). School officials typically regard the provision of such information as a necessary part of their accountability to the local taxpayer. And even though there are school districts that would rather not subscribe to such a means of accountability, they recognize that there is little choice. As the director of research in one school district explained, if he did not do the testing "the community would be suspicious because other school districts do it" (Interview, 12/18/79, p. 1).

Several additional reasons encourage a school district to select standardized norm-referenced tests. For example, some school officials see them as a helpful administrative tool, say for comparing student and staff performance in different schools and classrooms. Others may see them as a way to assist in the improvement of instruction.

For whatever reason or combination of reasons school districts might be in the market for a standardized norm-referenced test, they typically change tests anywhere from every three to every ten years. The decision to change tests is usually prompted by teachers' and administrators' dissatisfaction with the test in current use. Not atypically, a school official in one district explained that the superintendent had set up a committee to select a new test because teachers had done "a lot of grumbling and complaining about the test" (Interview with Research Coordinator, 2/1/80, p. 1). Such complaints may stem from the fact that the test used has fallen behind changes in the curriculum (Interview with Assistant Superintendent for Instruction, 5/23/80, p. 4), is "too long" (Interview with Director of Special Services, 2/5/80, p. 1), or is "not helpful to instruction" (Interview with teacher, 1/31/80, p. 3).

Despite dissatisfaction with a test in use, however, a new test is normally selected only if someone influential in the district presses for it. For example, a superintendent or head of the testing program may think that there is "a better test on the market" (Interview with Director of Pupil Services and Special Education, 5/23/80, p. 2)--that is, that a new test seems to address one or several complaints prevalent in the school district. On some occasions, a new test may be selected because it is reputed to be "easier" and promises to make a school district look better in the eyes of the community. For example, a superintendent in a small rural school district explained that he had decided on a new test mainly because his colleagues in another school district had told him that their "scores had gone up" (Interview, 12-13/79, p. 3). There was no evidence across school districts that this test uniformly produced that result;

some school districts using it found that their scores dropped. But school officials tried to select whichever test they thought would reflect favorably on their school district.

Choosing the standardized test to be used by the school district is not always easy. The view that a change should be made is rarely shared by everyone in the district and is usually balanced against other reasons for keeping the old test. These include the advantage of retaining the uniform and longitudinal data base provided by long-term use of the same test. For example, a guidance counselor spoke of his frustration when the board failed to approve his proposal to change tests because a board member did not want to lose the accumulated years of "baseline data" (Interview, 5/19/80, p. 1). Another common reason for not changing tests was given by a director of research. Though aware that the test in use was "out of date," he was reluctant to change, because "the new scores might go down," and this might bring the wrath of the school board and community upon him and the district (Interview, 2/5/80, p. 2). Another director of research was so concerned that the test scores in his district might drop if he changed tests that he piloted the new test before presenting his selection to the school board (Interview, 2/6/80, p. 3). Thus school districts decide to select a new test only if there are strong reasons for doing so; they are more often inclined to keep using the standardized norm-referenced test they had previously selected.

Who Is Involved in the Selection Process

Once the decision has been made to select a new test, most school districts set up a committee of representatives of test users. These committees usually comprise teachers, principals, central administration, and counselors. Generally a single person is in charge and guides the

selection process, because he or she is knowledgeable about testing--or at least is considered so. In large and medium-sized metropolitan school districts, this person is commonly the director or a member of an evaluation and research unit. In smaller and rural districts that have no evaluation and testing staff, the person chosen might hold a position in the area of Title I pupil services, counseling, curriculum development, or special education.

Whatever the formal role of the person responsible for selecting a test, he or she enjoys a wide degree of influence and discretionary power. For example, a principal who sat on the selection committee of his large city school district explained that he had relatively little control over selection because "there is only one person trained in the field here, and it's hard for anyone to compete with that knowledge." He added that the committee had been set up only because "the administration is committed to participatory management, but it is not a democratic system when it comes to testing . . . there is only one person who makes the decision and that's our local expert" (Interview, 2/9/80, p. 3). Those in charge of the selection process acknowledged their "local expert" status. As one such person commented, "Although I consulted with others informally, I pretty much made the decision" (Interview with General Supervisor, 3/13/80, p.5). But although there was one recognized or designated leader in the selection process, those in charge usually convassed the opinions and concerns of others in the district, either through the committee or by less formal means. As the head of an evaluation unit in a city school district explained, "/the districtwide committee/ will say I made the decision, and

I did, but I was also sensitive to them even if they did not know it" (Interview, 3/18/80, p. 2).

Those guiding the selection process tended to insulate the selection decision from the influence of private testing firms. For example, one director of research who was in charge of test selection explained: "Although I talked with sales people from private testing firms quite frequently before we decided to buy a test . . . my decision had much more to do with my own review of the material and comparisons to other tests on the market" (Interview, 2/6/80), p. 6). In other cases, the person in charge was careful to limit contact between testing firm representatives and those involved in selection. Rather than having representatives make presentations, for example, firms were usually asked "to send . . . their materials for review and to answer . . . questions" (Interview with Director of Pupil Services, 12/12/79, p. 13). This was due to a concern, particularly prevalent in smaller and more rural school districts, that testing firm representatives might sell them other than what they wanted. Private testing firms, on the other hand, were quite willing--some more than others--to send representatives to make formal presentations to local school districts.

The possible undue influence of private testing firms is generally of less concern to those in more metropolitan school districts, often because they have their own personnel knowledgeable in testing. Thus, it is more convenient for these districts to invite testing firm representatives to make formal presentations. But even here, the invitation to make a sales presentation is usually extended only after the test has been

reviewed by the person in charge, and less thoroughly by others involved in the selection. And for the most part those who select a test for district use do not regard private testing firm presentations as strongly influencing the decision. Rather, these firms' representatives were seen as simply answering the questions and responding to the concerns of those involved. As a member of one selection committee explained, the presentations just "provided us with information and pointed out the good features of the tests that helped us make a good decision." She added that as the presentations were all "outstanding," it was hard to say that they "had much influence" on the decision (Interview with Assistant Superintendent, 5/23/80, p. 3). One several-time participant in the local test selection process summarized this familiar process well when he said, "While most school districts are ostensibly at the mercy of private testing firms, they do their best to find out what they want to know before hand." Because of this, he continued, the role of most private testing firm representatives is usually "reactive" rather than one of hard sell (Interview with Director of Research and Evaluation, 2/6/80, p. 6).*

---

* The allocation of resources in the leading private testing firms suggests that they agree; the sales budget is considerably lower than that spent on test development. Even in the firm with the largest sales staff, the ratio between sales cost and development cost is 1:4. Thus these firms apparently attribute their ability to sell tests more to the reputation of the product than to salesmanship. There is wide variation, however, in the relative importance the major firms attribute to salesmanship. One company, for example, has a national staff of approximately 60, while another maintains a staff of five.

The consensus among those involved in district test selection, then, is that they make their own decisions and buy what they need, rather than what the testing firms want to sell them. Private testing firm representatives have little effect on those decisions because they are given little, if any, part in the selection process. More often than not, they are asked to make presentations to the school board or a group of teachers only after their test has been selected. Several school district personnel who take part in local selection have commented that the influence of private testing firm representatives pales in comparison with that of a textbook salesman. A general supervisor in charge of both curriculum and testing in one district explained, for example, that--in contrast to selecting a test--she had seen "a good textbook go by the wayside because the sales representative made a bad presentation, or a text . . . selected because somebody was very sophisticated and smooth in his presentation" (Interview, 3/13/80, p. 6).

If there is little evidence so far of circularity in the marketplace, it may nevertheless be occurring in subtler ways than through direct influence. Galbraith (1971), Fromm (1955), and others have argued that most private selling in this country is the result of sellers' shaping of consumer desires through mass communication (e.g., through advertising). And private testing firms do advertise, often through professional journals and flyers sent by mail. It would be difficult to show that these influences affect local decision making; but neither can it be shown that they are ineffective, or that circularity is clearly not at work in school districts' decisions to buy tests--though school officials

are probably not consciously aware of it.* Thus our tentative conclusion is what the available evidence suggests: little circularity is occurring in the educational testing marketplace, and school districts largely make their own decisions about the selection of a test.

But this still tells us little about the basis for their selection or their rationality as consumers. This remains an important question because, to the extent that they are rational consumers basing their choice on their needs and preferences, school districts can be seen as influential and powerful in the marketplace. The next section of the essay therefore focuses on the criteria by which school districts select a test.

THE BASIS OF SELECTION

The basis upon which a school district selects a standardized norm-referenced achievement test depends in part on that district's needs and circumstances. However, since district officials are commonly interested in selecting a test that will enable their school boards and communities to compare their district's student achievement with that of others across the nation, they initially select a number of candidate tests that seem likely to serve that purpose--that is, on the basis of technical quality. For example, the director of research in one school district explained that what went into his choice of candidate tests was their "technical quality." Similarly, an assistant superintendent of instruction said that the district's former director of research had

---

* This suggests that if circularity is occurring, it is impossible to determine what weight, if any, it should be given in this analysis.

shortened the committee's list of possible tests "simply on the basis of technical considerations" (Interview with Administrative Assistant to the Superintendent of Instruction, 5/23/80, p. 3).

The number of tests that are initially considered ranges from two to eight. Typically, the same four or five tests, provided by the leading major testing firms, are regarded as the technically better tests, and are therefore more seriously considered.* These four or five are usually considered to be of better technical quality in two ways, not mutually exclusive--their reputation, and certain methods employed for their construction.

Many school officials, especially in smaller and rural school districts, simply rely on the reputation of tests and their use by neighboring and metropolitan districts. For example, a director of curriculum in charge of the local test selection effort was asked how he knew that the three tests initially considered for adoption were technically reliable. He explained that they had been "statistically checked" by others. He added that, for him, selecting a test--at least initially--was like "buying a car--you ask other people whose opinion you respect and then you form an opinion of what is a good car" (Interview, 12/13/79, p. 4). In this case, he relied on some local experts in surrounding school districts to form his opinion of a technically sound test.

---

* These tests include the California Test of Basic Skills, the California Achievement Test, the Iowa Test of Basic Skills, the Metropolitan Achievement Test, and the Stanford Achievement Test, and are published by four testing firms: CTB/McGraw-Hill; The Riverside Publishing Company, a subsidiary of Houghton Mifflin; the Psychological Corporation, a subsidiary of Harcourt Brace Jovanovich; and Science Research Associates, a subsidiary of IBM.

Those in charge of the selection process in more metropolitan school districts typically relied less on reputation and presumed quality. Asked how they selected candidate tests, they pointed mainly to the norming characteristics of the leading handful of tests. These included how a test had been nationally normed--that is, whether there was evidence that the norming sample of students has been carefully chosen to represent the composition of the national population according to such variables as location, race, sex, and socioeconomic level; and how the content of the test was determined--that is, whether there was evidence that the test content was drawn from an extensive review of curricula used across the country so as to simulate a national curriculum as accurately as possible. A director of research conveyed this when he said that he checked whether tests he considered had "norming practices that were at least sound and /were/ going to provide a reasonable curriculum base of comparison" (Interview, 2/6/80, p. 4). Sometimes other construction features are taken into account. For example, an assistant superintendent in charge of test selection made sure that a test had been not only properly normed, but adequately "field tested," so that he knew it would work (Interview, 12/13/79, pp. 8, 9). But the preponderance of attention is usually given to their norming practices.

Typically, the norming practices of the leading four or five tests are checked without reference to published professional reviews of the tests. A director of curriculum who was in charge of local test selection, when asked whether his committee paid attention to such reviews, responded, "The hell with consumer reports." He added that in his district it was a matter of finding out what other people thought was good (Interview,

12/13/79, p. 4). This rather common attitude aside, outside professional reviews generally regard the same tests as indeed the better constructed (Buros, 1972; Center for the Study of Evaluation, 1975).

If this consensus may make the choice of candidates easy, however, it makes the choice among them difficult. Most school officials see the leading tests as technically similar and are unable to make any significant distinction among them. As one program evaluator put it, because "all the tests were pretty much the same as far as standardization was concerned," there was little to recommend one over another (Interview, 2/4/80, p. 2). Another local official explained that "all these instruments are about the same; they are all widely normed, and trying to make any determination of differences would have been a pain in the ass" (Interview with Director of Pupil Services, 12/12/79, p. 2).

Because these tests are technically rather similar, local school officials' problems with them are often the same. One common complaint is that the test content of the leading tests does not match the local curriculum, and that the tests therefore do not measure what is taught. One reading consultant, for example, found that because of the lack of fit between the local curriculum and that used by standardized tests, "if you try to place a child according to . . . achievement test scores with the materials /used/ for that grade level, the child would have lots of problems" (Interview, 5/28/80, p. 2). Similarly, an English teacher explained that the teachers in his school are "basically opposed to testing because it measures what they don't think they are teaching" (Interview, 1/31/80, p. 3).

Another frequent complaint is that these tests produce inconsistent results across grade levels. For example, a program evaluator had

serious reservations about using only one test because students seemed to score substantially lower with it at higher grades (Interview, 1/31/80, p. 2). Pointing to the same problem, the director of testing in another district explained that using a newer version of a test, students in the third, seventh, and eighth grades scored on average two to three grade equivalents higher than with the old version. Meanwhile, students in the other grades scored near grade level, as they had with the old version (Interview, 5/23/80, p. 2).

Often, those tests are also regarded as not relevant to the general level of achievement in the district. Students in a given school district, for example, might uniformly score off the bottom or the top of the test. A local school official in a district where most students typically scored at the bottom of the test called the test simply an "experience in frustration" for the students, while teeling teachers little about individual student achievement levels (Interview with Research Department member, 2/1/80, p. 5). In another school district, where most students typically score in the ninetieth percentile, the director of testing complained of the same problem--namely, that the test gives little indication of individual differences in student achievement (Interview with Director of Research and Testing, 2/5/80, p. 2).

The technical basis for these complaints is, for the most part, documented in the professional literature on testing. For example, it is often pointed out that standardized norm-referenced tests inadequately reflect curricular content and the objectives of schools and school districts around the country (Madaus, 1980; see also Porter, 1980;

Averch et al., 1972; Stake, 1972). Furthermore, critics charge that there is usually little correlation between standardized achievement test results and measures more directly related to instruction (Stadolsky, 1972; Madaus & Rippey, 1966).

The literature also points out that because standardized achievement tests are constructed to differentiate maximally among individual students, their accuracy in testing group performance and performance across groups is reduced (Carver, 1975; Porter & McDaniels, 1974). This might account for dissatisfaction with inconsistent comparisons of students in different grades. It is also argued that because standardized tests need to establish a normal test distribution, they cannot reliably measure achievement at the extremes of the distribution (Tyler, 1974)--which helps explain why officials in schools with mostly high or low achievers often complain that the tests provide little information on achievement differences.

Since the leading and better-constructed tests share some of these shortcomings, local school districts cannot easily choose among them on the basis of their norming practices. The most common problem that local school officials do try to address is that of the fit between a test and their curriculum. But this is difficult to do in a sound and systematic way--at least with a limited amount of time and money. As one local test expert explained, you simply "could not pick a test based on the curriculum . . . it has to be done on a trial and error basis . . . there is just not enough time and energy for someone to go through and do this in any sort of comprehensive way" (Interview with Director of Pupil Services and Special Education, 5/23/80, p. 5).

Not atypically, when comparisons were attempted they often turned up "very little difference" among tests (Interview with Assistant Superintendent, 5/23/80, p. 2). And professional reviews of these tests, even if they were conducted, would not offer much help, since they are concerned with the broader aspects of test construction rather than the particularities of school district curricula.

School districts, then, are rarely entirely pleased with the leading and technically superior tests; nor do they find it easy to select among them on the basis of their norming practice. One might thus argue that school districts are limited in their choice and their influence in the marketplace because they cannot buy exactly what they want. Yet school districts continue to buy these tests because district officials believe that they are basically sound and provide the best available nationally comparative information on student achievement. On this basis, one might argue that school districts do have some choice in the marketplace and have been able to influence private testing firms through those choices. If that were not so, one would expect little attention to be paid to the technical reputation of tests or to the use of certain norming practices in their construction; one would expect any test claiming to be standardized and norm-referenced to be considered on a par with the leading four or five.

Although the norming practices employed by the leading tests are similar, this does not mean that their tests are all the same. There is evidence that the tests differ in their ability to meet other school

district needs; and it is on this basis, unrelated to norming methods, that major testing firms compete with one another. An experienced hand in local test selection efforts summed this up well when he said, "Since there is such little difference in the norming practices with the large manufacturers of tests, it's the other stuff that becomes important" (Interview with Director of Research and Evaluation, 2/6/80, p. 6). And on this "other" basis, consumers see the major private testing firms as quite different and competitive.

Selection Considerations Unrelated to Norming Methods

Among the considerations unrelated to norming methods are a test's ability to cover a wide range of student achievement and its ability to provide information that is seen as helpful to instruction. Other test attributes include the length of administration and the appearance of the test. The supporting services provided by the test publisher, such as promptness in scoring and returning test results and the ability to provide needed technical assistance, may also be considered. Several of these attributes may enter into a district's selection decision; one or two of them may stand out because of a particular concern or bias of the local test expert. But local test selection decisions may well also represent, formally or informally, a composite of various considerations within a single school district.

Interestingly, one of the traditional criteria for consumer purchasing, the cost of the product, rarely has much influence in the selection decision. In part, this can be explained by the relatively uniform costs of the tests

put out by the major publishers, as well as by the relatively small percentage of school district budgets that is spent on testing programs. However, it is not uncommon for a local school district, having made a decision to buy a certain test, to negotiate for a lower price.

What follows is a discussion of some prominent considerations that may enter into the local selection decision--certain attributes of a test, and services provided by test publishers.

Test Attributes

As pointed out earlier, the leading four or five tests share common norming methods and this makes it difficult for school district consumers to select between them. Those selecting tests for school districts though often select that test which has been more recently normed. This is desirable since it promises to provide comparisons of student achievement with the most current national levels of achievement. Typically, this is a concern of the local test expert, usually an evaluator or member of the research department. Depending on his influence and the prominence of other local staff concerns, this may be the major reason why a certain test is selected. In one school district, the curriculum coordinator recalled that of the two tests they were considering, "one of them had older norms," and was therefore considered "no good" by the research department (Interview, 1/31/80, p. 2). Similarly, a math consultant in the department of curriculum reported that the committee selected the test they did mainly "because it represented the newest norm version of tests available" (Interview, 2/1/80, p. 1). If one of the leading tests was

recently renormed, it might thus be more favorably considered in a district seeking a new test. This was illustrated by a member of a research department in a metropolitan school district who explained why his test selection committee chose a particular test: "the other major manufacturers were just getting ready to put their new tests on the market, but this one had already been renormed and was in operation--so we went with it." (Interview, 2/1/80, p.2). Because private testing firms recognize this, they compete to put newly normed and updated tests on the market. This cannot be done too frequently, however, because the development and construction of a standardized norm-referenced test is estimated to take approximately four years (Interview with Director of Research at CTB/McGraw-Hill, 2/7/80, p. 3).

When during the school year a test was empirically normed may also be an important consideration to the local test expert. This is because the selection of a test normed at the time or times of the year when local testing occurs promises to provide more accurate comparisons with national norms. As one local test expert, a director of special services, recalled of his tenure on the test selection committee, a test was chosen that had "two . . . norming dates, and this compared favorably with most other standardized tests, which only had one norming time." This was seen as desirable since it gave the school district greater flexibility in deciding when to test, without having to sacrifice any accuracy in test results by testing at a time different from that at which the national norms were determined (Interview, 2/5/80, p. 1). Typically, this consideration was important in larger school districts with evaluation

and testing specialists who were interested in maximizing the accuracy of district test results. In school districts without a test specialist, the norming of tests was accorded little importance and rarely entered into selection decisions.

Another attractive test attribute is the ability of a single subtest of a K-12 battery of achievement tests to measure across two or three grade levels. This allows students greater opportunity to be tested at their level of achievement rather than simply at their grade placement, and is especially valuable where a student's achievement level is several grades higher or lower than those of his or her cohort in the same grade. For example, a program evaluator explained that one reason for selecting one of two leading tests was that it had "multi-level testing." This was a particularly important consideration for his school district because the students generally scored below grade level and "they got crazy results when they had to test them only on one grade level." By crazy he meant that most students scored uniformly at the bottom (Interview, 1/31/80, p.1).

The multi-level capability of a test were particularly attractive to administrators and teachers. Administrators liked it because it simplified the test administration; one test could be given for several grade levels. Teachers often liked it because it allowed their less advanced students to answer questions at a lower grade level, and therefore not to feel overwhelmed and frustrated by the test.

One of the leading tests may also be selected because of other attributes, such as ability to aid in classroom instruction. A test may, for example, allow the scoring of groups of students in terms of mastery of certain skills. Thus the director of test selection in one

district sought a test that "teachers would use"; otherwise, there was no purpose in testing. With this in mind, he explained, one of the reasons he selected the test he did was teachers' response to a survey he conducted. The teachers liked a particular test because it had "a profile breakdown--that is, it related certain items to the instructional objectives identified." According to him, the profile breakdown would enable teachers to see how a class performed on the test according to various skill objectives and so help them identify what areas to stress in their teaching (Interview with Director of Research and Evaluation, 2/6/80, p. 2). A related feature that some considered important in promoting instructional use was a test's ability to provide specific diagnostic and prescriptive information. For example, an instructional specialist in reading explained that she and the teachers on the selection committee wanted a test offering the type of feedback provided by the mini-tests at the back of some reading series--on areas in which the students were weak, and on what part of the test they could practice to improve. The committee therefore selected the standardized norm-referenced test that provided the most detailed data on individual student performance, in the hope that it would be useful for "remedial" instruction (Interview, 3/13/80, p. 2). The test selected in this case provided references to several major textbooks that corresponded to the areas of weakness indicated by the test results. And in some districts, administrators--both principals and central office staff--hoped that using a test capable of improving instruction would result in students' scores going up--and that the school staff then would look good in the eyes of the community. Teachers, however, were typically more interested in the usefulness of test results for classroom instruction.

Administrators and teachers were also generally attracted to another test attribute: the provision of student achievement scores in comparison to an "ability" or "anticipated" score (the latter being based on an intelligence test contained within the achievement test). This score is usually seen as potentially useful in teachers' instructional endeavors. One principal called it a "helpful tool" to teachers, which could serve as a "validity check" to determine whether a student was doing what he or she was capable of (Interview, 5/14/80, p. 2). Similarly, a teacher explained that he and his colleagues on the testing committee were attracted to the "anticipated score" because it enabled teachers to "really get down and help the student" (Interview, 12/18/79, p. 7). Until recently this feature was offered by only one of the major private testing firms; and because it can be an aid to instruction, it was influential in a large number of school district selection decisions.

Other considerations that may enter into local selection include the time required to administer a test, or even its appearance. Of these, test length was the most frequently mentioned. For example, as a member of one selection committee explained, most of the teachers wanted a test similar to but shorter than the one they were using (Interview with Director of Special Services, 2/5/80, p. 1). Similarly, one reason why a principal pushed for the test eventually adopted was that it could "get in and get out and did not take long for kids to do" (Interview with Principal, 5/14/80, p. 1). Moreover, particularly in the lower grades, teachers "did not want to have their kids sit for so many hours taking /the test/" (Interview with Research Coordinator, 2/1/80, p. 2).

As these examples indicate, the general feeling was that the shorter the test, the better. Typically, this was seen as important by teachers and principals who often felt that testing took them away from providing more direct instructional services.

The appearance of a test was sometimes mentioned by teachers as a consideration in selection. An assistant superintendent of instruction said that one of the subjects discussed in their committee, composed of teachers, was the format of the test and whether it was "in red, blue, or green." She added that this sounded rather simplistic, but "was important because it was difficult, for example, to read certain tests which were printed in red . . . on white paper" (Interview, 5/23/80, pp. 22-23). A member of another test selection committee explained that one of the "hot issues" that emerged was the size of the print. He commented that it was "funny" that this should "/end/ up playing a role in the decision," that that it was apparently one of the reasons teachers did not like the old test (Interview with Administrative Assistant, 5/23/80, p. 5). In smaller and more rural school districts, often without professional testing and evaluation staff, such test attributes as a short time of administration and an easy-to-read format played an important role in the selection decision--usually made by principals and teachers.

Service by Test Publishers

Another consideration in the selection decision, generally less important than those discussed above, is the kind and quality of services that the major testing firms provide with their tests. One such service, usually of more interest to administrators and evaluators than to teachers,

is test manufacturers' prompt return of test scores to school districts. For example, a director of curriculum, who had sat on several test selection committees, called the "quick turn-around time for a test" a common concern (Interview, 2/4/80, p. 2), because it allowed the placement of students at the beginning of the school year, or the checking of their achievement at its end. Thus, when a new test was selected the testing firm that could demonstrate a potentially short turn-around time usually held a competitive advantage. In districts that scored their own test or made scoring arrangements independently of the test publisher, of course, this consideration did not enter into selection decisions.

District personnel also commonly mentioned the availability of technical assistance or "consultant services" from private testing firms as a consideration. For example, the director of testing in a large city school district wanted to find a testing firm whose representative he could call "and have him come down to help me put in a new order for next year or help me sort out some problems with the testing programs." He added that he was simply looking for a testing firm that followed good business practice--"not to sell and run" (Interview, 12/18/79, pp. 8-9). In some other school districts, the availability of technical assistance for such matters as test interpretation, workshops for teachers, and presentations to the board also could influence test selection, though commonly it was secondary to other considerations.

The range of considerations on which school districts ultimately select among the leading four or five tests is thus quite wide, and varies greatly across and within school districts. Even within a single school district, different people involved in selection decisions may agree on

a given test for different reasons. For example, in one school district central administrators and Title I personnel were interested in a test primarily because its publisher promised to score and return it promptly; and principals and teachers tended toward the same test because it was short and easy to administer. In this example, the test was selected because it combined features that were attractive to the different staff involved in selection. In other cases, those involved in a selection decision shared a primary interest in certain test attributes and selected a test on that basis. For example, a school district with a strong commitment to using test results for improving instruction was particularly attracted to that norm-referenced test which they believed provided the best breakdown of student achievement according to specific skills.

Through this process of sorting and weighing the various features and capabilities the leading tests provide, districts thus settle on the test that comes closest to meeting their usually fragmented needs and wants. Examination of that process made it clear that the leading private testing firms provide different options and services to meet those needs and wants. Typically, those that offer the more attractive options and services enjoy the business of more school districts. Indeed, a review of the past tests of these firms shows that they have been changed in numerous ways so as to make them more appealing and more responsive to consumers' needs. Efforts have been made to ensure that the tests are relevant to and useful for instruction. One private testing firm in particular has added several new features to its tests, and has been successful in capturing the lion's share of the market.

However well they may meet the internal requirements of school districts, the purchasing of tests from private testing firms is also

influenced by state and federal requirements. The nature and extent of that influence is discussed below.

State and Federal Testing Requirements

Clearly, the stronger influence on local test selection is that of state's testing requirements that mandate the use of a particular norm-referenced test for certain grades.* This influence is strong because local school districts, with few exceptions, use the same test for local testing in other grades. As an assistant superintendent explained, his district decided to select a particular test for grades K-12 "basically" because "the state had selected a test" and mandated its use in grades 4, 8, and 11 (Interview, 3/14/80, p. 1). Similarly, a district director of testing said that the state mandate of a norm-referenced achievement test was the determining factor in their decision: since "we were using it to meet the state mandate already, it seemed the natural thing to stick with it" (Interview, 2/4/80, p. 1).

A state-mandated test is influential for several reasons. One is that the test results are often used by the public, if not the state, as a way of evaluating local school districts. Thus, it makes sense for a school district to use the same test as the state, so that any curricular or instructional weaknesses can be identified and addressed. A second

* States may also be influential in local selection decisions by providing a list of approved tests, as is the case in New York and Wisconsin. Typically, this list includes the same tests that are ordinarily considered to be of superior technical quality by local school districts. Though no districts in states with such lists were visited in this study, it is likely that their selection is based also on the considerations discussed in this essay.

reason is that local school officials, especially in smaller districts, are open to the influence of the state since it diminishes their responsibility for selecting a test. The major private testing firms are aware of this, and consequently there is much competition among them for state selection of their tests. Winning a state nomination greatly increases the likelihood that school districts in that state will select the same test for their testing programs.

In mandating a certain test, however, the state may be responding in part to local preference. For example, an associate director of testing who was in charge of the state's test selection said that they had chosen a particular test because "this test and testing firm was familiar to local school districts and had already done a very good job" (Interview, 3/10/80, p. 2). State officials also are influenced by local officials who sit on the state selection committee, and who more often than not are drawn from a large city or cities and are knowledgeable and politically powerful. Thus school districts have considerable influence on test choice at the state level.

Federal evaluation requirements, especially for Title I, can also enter into school districts' test selection. But typically, because no one test is mandated by the federal government, that influence is indirect. School districts may decide to select for local testing programs the same test that is used for evaluating federal programs such as Title I. This enables them to use a single test for their major testing needs rather than having to test students enrolled in Title I with a different test than that used for other students. For example, in one such school system

a member of the test selection committee, asked about their basis for selection, replied: "The Title I people seemed to have some rather firm ideas about what they needed," and added, "They had certain constraints we had to operate under" (Interview with Coordinator of Curriculum K-12, 3/31/80, p. 1). The constraints imposed by testing for Title I programs often include an effort to match the norming dates of tests and those of school districts. For example, an assistant director of evaluation in a large school district pushed for one test, which was eventually adopted, because "it had dual norming dates and was easier to fit into /Title I/ guidelines which required testing within three weeks of the norming date" (Interview, 2/14/80, p. 1). Similarly, a member of a test selection committee in a large city school district explained that they had previously selected a test partly because it "was normed in the fall and spring and therefore could be used for Title I pre- and post-testing" (Interview with Research Coordinator, 2/1/80, p. 2). And an evaluator for special projects, including Title I, wanted to make sure that the test selected by his committee had a "diagnostic component," so that the program people could use the results for providing remedial instruction (Interview with Coordinator of Special Projects, 2/6/80, p. 2).

In those cases, then, where school district testing was merged with testing for Title I, the particular needs of Title I testing played a part in the selection of a test for local programs. Private testing firms, aware that a test meeting Title I requirements may also be used for local testing programs, thus compete quite intensively for the Title I market.

While federal--especially Title I--and state testing requirements can be influential in local test selection, school districts control the extent of that influence. Even state selection of mandated norm-referenced tests is based partly on previous local selections and on political factors. Furthermore, school districts need not use the state mandated test for their own testing if they consider it inappropriate. Similarly, the influence of Title I testing requirements is limited to local districts' selection of tests for Title I evaluation, unless districts decide to use the same test for their own testing. And even then, district personnel can select that test which best meets some combination of their Title I and district testing needs. Thus school districts, while influenced directly and indirectly by state and federal requirements, are still able to exercise their preference as consumers in the marketplace.

THE INFLUENCE OF SCHOOL DISTRICTS AS CONSUMERS

This account of how and on what basis local school districts go about selecting educational achievement tests suggests several observations about their influence as consumers. On the one hand, school districts can and do exercise choice in the marketplace, and so are able to influence the tests produced by private testing firms. Their power as consumers is shown by the fact that these firms provide technically well-constructed tests--at least as far as certain norming practices and procedures are concerned--that enable them to compare the achievement of their own students with that of the larger national population. And even though this selection criterion has seemingly limited the number of testing firms to a handful that share a virtual monopoly over the norming

of such tests, there is much evidence that districts select among the leading tests on such grounds as their ability to improve classroom instruction, to simplify the administration of testing, and to meet state or federal testing requirements, to mention a few. That such choices are offered indicates that school district needs influence the kinds of tests produced. Otherwise, one would expect tests to differ very little from one another, and districts to take any test at random, having no basis for selecting one over another. Similarly, one would not expect private testing firms to compete with one another at all.

On the other hand, this account suggests that while school districts enjoy a good deal of choice and influence in determining some aspects of the tests they buy, they are limited in their influence over and satisfaction with the way tests are normed. They cannot know how accurate the information is that tests provide on comparative student and local school achievement, or whether it does, in fact, meet their needs. For example, they cannot systematically determine whether and to what extent the tests developed by the major private testing firms take into account what they are teaching their students in their schools and classrooms. Indeed, there is some evidence that most local school practitioners, especially teachers, think that these tests bear little relation to what is taught in their classrooms; at best, they see the tests as broadly and vaguely related (Kennedy, Apling, & Neuman, 1980). Similarly, school districts are unable to know or determine to what extent norming samples of these tests are representative of students in their district. The

student population of most districts rarely matches that of a stratified random sample of the national student population used by private testing firms. Nevertheless, the achievement of their students is compared to that of the sample population. With respect to test construction, then, school districts' influence over private testing firms is limited.

This discussion thus far has dealt with the choices school districts have and their ability to influence private testing firms by exercising those choices. And it has pointed out that, to the extent that they cannot exercise choice, their ability to have testing firms produce what they want, and to hold them accountable, is limited.

It has been assumed throughout that school districts are rational consumers who know what they want and express their preferences through choices in the marketplace. But perhaps school districts are less than rational consumers. For various reasons, districts may be unclear or inconsistent about what they want. They want, for example, to compare local students' educational achievement to that of students nationwide--and yet want a test that will reflect the local curriculum and student population. And the two are inherently incompatible, since there is no national curriculum or standard student population upon which a test can be based.

School districts, then, may be shopping for what does not exist. It is not that they are simply asking for too much--apples and oranges to be put in the same basket--but that they want what cannot be provided--fire and ice in the same container. It may be in the nature of the local education enterprise that they are bound to want both--to see what their

own students are learning from what is taught, and to compare this objectively to what other students are learning and are being taught. But since school districts have different curricula and are composed of different student populations, school districts cannot have both. If this is indeed the case, school districts could not get what they want from these tests even if they exercised more control in the marketplace.

THE MARKETPLACE AS A MEANS OF LOCAL CONTROL

Given the nature of the local educational enterprise and its not wholly rational wants, it is unclear whether the marketplace can work better. School districts want a test that can standardize what are essentially local conditions so that an objective comparison of student achievement can be provided. Private testing firms have responded to this demand; given its irrational nature, it is difficult to see in what ways they could be still more responsive. For example, as said earlier, no educational achievement test can be constructed that matches both the curriculum of any single local school district and the curricula used in all other school districts around the country. Private testing firms already make an extended effort to take account of the various curricula, texts, and instructional techniques of different school districts. In fact, evidence of this endeavor in standardized norm-referenced achievement tests is one criterion upon which school districts base their selection of a test in the first place. Furthermore, where school districts are clear about the intended use of a test (e.g., for improving classroom instruction, meeting state and federal requirements, or administrative accountability), they can select a test containing those features that best meet their needs. And school districts

who know they want a test that is relevant to and compatible with their curriculum can additionally buy criterion- and objective-referenced tests from the same or other private testing firms. Indeed, the fact that school districts have expressed a need for this type of test (Rozczinski, 1977; Ebel, 1974) has contributed to a rapid increase of new testing firms and tests on the market.

The marketplace as a means of local control, then, may work insofar as consumers know what they want and can thereby hold private testing firms accountable. Where school districts find that their needs and wants are not met through the marketplace, this may have less to do with their power--or lack of it--as consumers than with the rationality of these needs. Private testing firm development and provision of standardized norm-referenced achievement tests may simply reflect mixed or inconsistent messages from school districts.

Private testing firms operating in the marketplace may thus well be the best vehicle for maximizing local control. They play a unique, if imperfect, role in taking account of local educational conditions and at the same time providing a neutral base for assessing national student achievement. It is not clear that there are alternative governance arrangements that could better provide this delicate balance. But before assuming such a conclusion, it is best to consider some possible arrangements.

THE MARKETPLACE VS FORMAL GOVERNANCE ARRANGEMENTS

One alternative is for local school boards or their representatives to develop their own tests. This would have the advantage that the test would be based on their own curriculum and student population. But there are several problems with this. First, districts would likely be unable to provide much of a comparative base for assessing their student achievement vis-a-vis that of other school districts in the country. In addition, they probably lack the technical ability, manpower, and other resources to provide a technically credible tool of assessment, and may thus require outside expertise. This may be particularly the case with smaller school districts that do not have the staff for in-house test development efforts. Furthermore, test development by local school boards and their representatives may involve extended political debate and conflict with others in the district over both substantive and methodological issues. This has already been a problem with some large school district efforts to develop their own objective- and criterion-referenced tests (Interview with Director of Research, 1/2/81, pp. 2-4; Interview with Program Evaluator in another school district, 1/1/81, p. 3).

Current market arrangements avoid many of these problems. Private testing firms, by virtue of their national role, have access to a wide variety of school districts and students for comparing achievement and can develop tools of assessment at relatively low cost because of the large number of consumers. They also offer an independent and authoritative base for comparisons among school districts and thereby minimize political and public debate at the local level.

Another alternative to the existing market arrangement is for states to assume more responsibility for developing educational achievement tests. An advantage of this approach is that state education agencies are possibly more in touch with and sensitive to local school district curricula and student populations than are national organizations like private testing firms. In addition, state agencies are formally accountable to elected officials, so that school districts can lobby for changes in the way tests are constructed. Furthermore, many state agencies have the necessary resources for developing such tests.

A major disadvantage of this alternative is that state agencies would be unlikely to be able to develop a standard test that can provide much of a national comparison. More likely, each state would be inclined to develop its own state test, rather than national achievement tests, as has already been done, for example, in California and Texas (Ebel, 1974; Pipho, 1978). Additionally, these tests are likely to be geared primarily to meeting state needs--such as accountability--rather than those of local school districts. There is already evidence of this occurring in California and Texas school districts' reactions to the state-developed tests; they typically buy standardized norm-referenced tests as well, to better meet local needs. For example, as one program evaluator in an affluent Texas school district explained, in addition to administering the state mandated test, he also bought a standardized norm-referenced test for his district because "the state assessment program was not very helpful in comparing /our/ students to those in

the rest of the nation." He added that after all, students in his district would be competing against the national population for the rest of their lives, so why compare them only to students in Texas now? (Interview with Program Evaluator, 1/31/80, p. 1). An evaluator in another state with its own assessment program reported that while the state test created lots of media interest "because districts across California can compare themselves in reading and math, it is not very valuable to teachers." He explained further that, as a result, only a few even looked at the scores. His view was that standardized norm-referenced tests were much more helpful to teachers because of their analysis of individual students' achievement--or lack of it--in the same areas (Interview, 8/19/81, p. 2). State-developed tests are unlikely to be as useful to local school districts.

A related disadvantage of states developing their own tests is that these may be seen as a threat to local school district control. Through their testing program, states may be exercising more control over local curricula than local school districts want.* This may lead to extended and complicated debate between state and local agencies over issues of assessment. In some cases the validity of the state test has already been challenged in the courts (e.g., Debra P. v. Turlington in Florida). Local school districts are far more receptive to tests produced by private testing firms because of their political

* Madaus (1979), for example, discusses this as a possible consequence of state minimum competency testing programs.

neutrality and objectivity--despite apparent shortcomings in their products.

Again, there are some telling disadvantages in having state agencies take a more active role in providing tests than under current market arrangements. One is that state agencies would not be able to provide much of a national base for comparing student achievement. Another is that state-developed tests are likely to be designed to meet states' own needs and would not easily allow for the exercise of greater political control by local school districts. In addition, state-developed tests might create another arena for conflict between local and state agencies. For these reasons, school districts are likely to want to purchase tests from private testing firms anyway. This appears to be the case in those states that have already developed their own tests.

It is difficult, then, to assume that formally responsible government agencies, at the local or state level, would better serve the needs of school districts for educational assessment than does the marketplace. A further alternative is for the federal government to develop a test for use by local school districts, on either a mandatory or a voluntary basis--as has already been proposed in Congressional hearings (Shoemaker, 1978). The major advantage of this alternative is that it would offer a national base of comparison for assessing student achievement. But it suffers from a significant defect--one that became clear during the course of these hearings: the federal government's development of such a test might lead to an undue exercise of control over local school districts and be inconsistent with Constitutional delegation

of authority over education to the state. This is indeed a major problem.

Another possible role for the federal government, one more consistent with its Constitutional role, would be to regulate the development of national tests in the marketplace. This might be done, as in other areas (e.g., air travel), through some form of national regulatory agency (Hobbs, 1975), which would develop standards for tests and approve some tests over others. For example, it might provide some check on the curriculum basis and morning samples of the tests designed by private testing firms. It might also provide consumer reports to disseminate its findings to the public.

School district political control over such an agency, however, would likely be problematic. The needs and preferences of districts would differ and compete with one another for influence. The agency would therefore be constrained in responding to any particular district and would likely become a lightning rod for many districts' criticisms and complaints. For political survival, then, it might have to confine its regulation of private testing firms to technical issues of test construction. It would thus probably have no more power and represent school district interests no better than existing private professional organizations like the APA, or NCME, or AERA, that already issue technical standards for test construction.

In addition, a regulatory agency might be open to the influence of private testing firms acting in their own behalf. There is considerable

evidence that this occurs in other areas of government regulation of private industry (McConnel, 1967; Bauer, 1963). The result might be the legitimization and ossification of the most powerful existing private interests in the marketplace, as some have claimed has happened under other federal agencies (e.g., the Civil Aeronautics Board /Hector, 1960/; the Food and Drug Administration /Hunter, 1972/; and the Federal Communication Commission /Friendly, 1959/). Thereby, private competition might be restricted and its incentive to meet local district testing needs and preferences reduced. Conversely, the agency might end up interfering with the activities of private testing firms and hinder their ability to respond to consumer needs. A common result of such regulatory agencies in other areas, for example, is to create additional administrative costs for private industry, and in turn higher consumer costs (Scherer, 1971; Thurow, 1980).

SUMMARY AND CONCLUSIONS

In summary, then, while formal governance arrangements do offer local school districts more political control over the development or regulation of tests at the local, the state, or the national level, they also suffer from two major disadvantages compared to the marketplace. One disadvantage is that the formally constituted role of governance of local, state, and federal agencies limits the extent to which they are legitimately responsible for developing national educational achievement tests. For example, because states and local school districts are limited to their geographic areas, it is unlikely that they could publicly justify the necessary resources for the

development of a national test. The federal government does have some responsibility for education at the national level. But since formal authority over education resides at the state and local levels, its national role would be constrained in either the development or the regulation of a national test. In contrast, private testing firms are not constrained by formal political boundaries; their role as defined by the marketplace is national in scope.

A second reason why local school districts are unlikely to prefer formal governance arrangements is that they would invite political conflict over the methodological and substantive issues of assessment. Local officials, for example, may well find themselves involved in assessment issues at the local, state, and national level--hardly a welcome addition to the already overwhelming demands placed upon them. Again, this problem is avoided in the marketplace where private testing firms are without formal political means of control or accountability; instead, they are dependent upon their consumers' informal perception of them as a neutral source of authority.

Formal government agencies, then, in assuming more control, are limited in the extent to which they can claim formal responsibility for developing a national test of educational achievement; and they run the risk of becoming embroiled in political conflicts over the test development process. Furthermore, even if they were able to avoid these pitfalls, it is not clear that more control would necessarily lead to the development of a better test. There are inherent limitations in the ability of politics to produce the desired technical outcomes in

the case of a national test of educational achievement. For example, the undue influence of a group of school districts over the curriculum base of a state-developed or nationally referenced test would likely compromise the standardization of its comparative base, and in turn the accuracy of its results. Greater political control by school districts may not, in fact, provide them with what they want. There may simply be an inverse relationship between maximum school district control and technically better tests.

The limitations of politics for producing desired technical outcomes in the area of testing are perhaps analogous to, although not identical with, other areas of formal government involvement in public policy areas. For example, while organized local, state, and national public interests may be able to influence the Environmental Protection Agency and its car pollution regulations, car manufacturers may simply be unable to produce a more pollution-free engine while providing the public with a desired level of automotive power. Similarly, consumer interest group pressure on the Food and Drug Administration to take preservatives out of food because of possible harmful side effects may work against other consumers' preference for long-lasting food. As in the case of testing, the ability to produce a product that will meet all consumer needs may be constrained by technical limitations--despite the opportunity for the political expression of these needs through formal governance arrangements. Government involvement may simply provide an arena for public and political debate while contributing little to, if not compromising, the meeting of public needs and wants through the marketplace.

In conclusion, then, it appears that the major strength of formal governance alternatives to the marketplace--i.e., providing local school districts with more political control--may actually militate against the satisfaction of school districts' needs. Private testing firms, operating through the marketplace without formal political control, seem better able to meet those needs. This is a curious conclusion in some ways, since it suggests that the exercise of school districts' political control through some form of formal governance arrangements is less effective than their exercise of control and influence through the marketplace. The marketplace, at least in the case of achievement testing, may provide a degree of public control that formal public decision making and politics cannot.

METHODOLOGY

The purpose of this study was to explore whether and to what degree private testing firms (which publish norm-referenced educational achievement tests) are held accountable by local school districts. It was assumed that local districts' primary means of holding private testing firms accountable is through their role as consumers in the marketplace. With this assumption in mind, the study was designed to gain an understanding of how and on what basis local schools decide to buy norm-referenced achievement tests from private testing firms.

Little previous work exists to guide the study. What related research is available focuses on particular aspects of the test selection process--e.g., teacher involvement (Ward, 1980), or determination of the information needs of the test selectors. It also relies exclusively on survey methods and provides little in-depth analysis of local test selection. Furthermore, little, if any, attention has been paid to local school district views of and relationships with private testing firms in the selection process. A field-based case study method was therefore adopted because of its established appropriateness for providing detailed and encompassing views of such uncharted areas of inquiry (Bogdan, 1975; Stake, 1977; Glazer & Strauss, 1973). What follows is a description of the method of this study. It is divided into three sections: site selection, conduct of field research, and recording and analysis of data.

SITE SELECTION

In constructing the sample for this study, an initial effort was made to select local school districts that were nationally representative of those who buy norm-referenced achievement tests from private testing firms. Random selection procedures constitute the traditional approach to defining any such sample. Under this approach, probability theory can then be used to generalize the results obtained to the universe from which the sample was drawn. However, random sampling was not appropriate in this case, since the sample selected through this means would likely be considerably larger than could be accommodated by the resources of the study. Instead, an effort was made to select local school districts according to the factors that may help explain their test selection decisions. Two such factors were identified through preliminary interviews with people in the field: district size and state testing policies. (Other factors considered included geographic location and per-pupil expenditures of local school districts; however, as there was little confirmation that these were important, they were not used in site selection.) A third factor that was taken into account was the ease of entry and access to school districts. The grounds for the two substantive criteria are discussed below.

Selection by State

Several considerations went into the selection of states in which local sites were chosen. These considerations stemmed from the recognition, gained from both the literature and interviews with people in the field, that certain state testing policies may influence local school districts' selection of tests from private testing firms. While state test selection policies vary considerably across states, two policies in particular were identified as important to include in the sample. One of these was state policy mandating the use of a commercially published norm-referenced achievement test. California was selected because of its mandated use of a state-developed criterion-referenced test. And Massachusetts and Connecticut were selected because of the virtual absence of policy requiring a particular test to be used.* In each state, five local school sites were visited; an extra pilot site was visited in Massachusetts.

* Massachusetts has recently implemented a minimal competency testing program; but it allows school districts to use either a state-developed criterion-referenced test on a commercially published norm-referenced achievement test.

Selection by Size

Twenty-one local school sites were visited for the study. In selecting these sites, an effort was made to represent school districts of different size. To that end, three categories of school district size by student enrollment were developed from the seven standard categories used by the National Center for Educational Statistics: "large school districts," with enrollments of over 25,000; "medium-sized school districts," with enrollments of 4,000 to 24,999; and "small school districts," with enrollments of 4,000 or less. Sites in each category were then selected with two considerations in mind. One consideration was to change sites in proportion to the N.C.E.S. breakdown of local school districts by size. (This meant that 1.2% of the sample should be selected from large districts, 10.3% from medium-sized districts, and 89.5% from small districts._ But this consideration alone seemed inadequate, since the literature (National Association of School Boards, 1977; Center for the Study of Evaluation, 1978) and preliminary conversations with people in the field suggested that only a low percentage of small school districts bought commercially published norm-referenced achievement tests. Additionally, it was pointed out that the selection of such tests by a larger school district might well prove to be influential in the selection decisions of smaller districts. Therefore, it was decided to increase the proportion of sites selected from among the medium and large districts. Hence 5 sites were selected from among the large districts--representing 24% (approximately) of the sample; 7 sites were selected from among the medium-sized districts--representing 33.5% (approximately) of the sample;

and 9 sites were selected from among the small districts--representing 43.5% (approximately) of the sample.

CONDUCT OF THE FIELD RESEARCH

The field research was conducted in three different settings: state educational agencies, private testing firms, and local school systems. However, the preponderance of the field work was conducted in local school districts, since the study was primarily concerned with their selection of tests. The field research in each of these settings is discussed below.

State Educational Agencies

In each of the states selected for the study, site visits were first made to state departments of education. Relevant state officials were interviewed, and documents regarding state testing programs (or proposed programs) and activities were gathered. Suggestions were also sought regarding appropriate local sites to visit. Ten state department officials in all were contacted and interviewed.

Private Testing Firms

Representatives of the major private testing firms were interviewed either personally or on the phone throughout the course of the field research. Those interviewed were selected because of their marketing background and/or direct contact with local school districts. Interviews were held primarily as a check on data gathered from school district officials as well as to explore particular issues that emerged from visits to school districts. Six such interviews were conducted,

which typically confirmed what was reported by school district officials. Documents of a technical and public relations nature were also gathered from these contacts with private testing firms.

Local School Systems

Two methods were used to gather data at the local school district level--interviews and document review. Of these, interviews with school personnel were the primary means. The format for interviews was semi-structured and open-ended and was guided by a protocol developed according to the issues under study. Interviews generally ranged from 45 minutes to an hour, though some were as long as an hour and a half, while others took only one-half hour.

Who was interviewed in each local school system depended upon who was involved in test selection or was knowledgeable about it. This varied according to the nature and complexity of the test selection process in each system. Enough people were interviewed in each system to provide a clear and thorough understanding of that district's selection process. Typically, this involved several interviews within each site, with members of the evaluation and testing department, district administrators and supervisors (generally in the math and English curriculum areas), school principals, guidance counselors, Title I personnel, and teachers. The total number of interviews conducted in local sites was 104--approximately five interviews per site.

In each site, several documents were reviewed and analyzed. These documents primarily comprised internal school district memos, usually

pertaining to the composition of school district selection committees, and minutes or summaries of committee meetings. Documents reviewed also included information provided by private testing firms to local school districts. This information ranged from technical booklets to public relations literature. Where relevant, other documents, such as the list of district curriculum objectives, were also reviewed. Document review typically occurred on site so that school personnel involved in the selection process could be identified and interviewed, and issues pertinent to local decisions could be pursued during the visit.

RECORDING AND ANALYSIS OF DATA

Field Notes

All interviews and conversations, both formal and informal, were recorded in field notes. Notes were taken by hand during interviews and conversations and later dictated for transcription. They were discursive in form and frequently wide-ranging. People were often quoted verbatim and the notes became the researcher's personal record of what was said. An interview of one hour produced field notes which, when transcribed, had an average length of 6 to 8 single-spaced pages. In addition to recording interviews, the field researcher dictated summary notes on site visits and notes describing program and district context. Close to 400 pages of field notes were transcribed.

Data Analysis

After field notes were transcribed, preliminary analyses were conducted and recorded in memos. These memos were submitted to the senior consultant to the study for review. New areas of inquiry were suggested,

and different ways of pursuing issues identified earlier were discussed. This provided a basis for further developing and refining of the interview protocol used in the field work. A reiterative process of field work and analysis was thereby established from the early stages of the study. Thus by the end of the field work several memos representing progressive analysis of the data collected in the field were available for review. These analyses provided the map for coding the field note data and for the final analysis. One of the significant results of this process was that, while considerable diversity in school district selection processes was recorded, several common and general themes emerged when the data were analyzed across sites. Thus it was decided to present the findings for the study in cross-site format rather than in individual cases. In writing the final report, a concerted effort was thus made to select illustrative examples from the sites visited that best illuminated these themes.

Additional Data and Analysis

During the course of the study, it became apparent that an NIE-funded research project concurrently being conducted by Huron on local school districts' use of evaluation and testing information was collecting data relevant to some of the issues pursued in this study. Toward the end of the study, an effort was therefore made to cull and review the relevant data from the field notes of that study. This allowed data from 18 additional sites to be included in the final analysis. Where appropriate, illustrative examples from these sites were also used in preparation of the final report.

REFERENCES

Association of American Publishers. Estimated Book Publishing Industry Sales. Appendix, Table 51. New York, 1980.

Averch, H., Carroll, S. J., Donaldson, T. S., Kiesling, H., & Pincus, J. How effective is schooling? A critical review and synthesis of research findings. Santa Monica, Calif.: Rand Corporation, 1972.

Bachrach, P. The Theory of Democratic Elitism. Boston: Little Brown, 1967.

Bailey, S.K., & Mosher, E. K. E.S.E.A.: The Office of Educational Administers a Law. Syracuse: Syracuse University Press, 1968.

Bauer, R., Paul, I., Dexter, A. American Business and Public Policy. New York: Atherton, 1963.

Bendix, S., & Lipset, M. Class, Status, and Power. Glencoe Free Press, 1963.

Bogdan, R., & Taylor, S. Introduction to Qualitative Methods. New York: John McKay & Sons, 1975.

Boston Globe, November 12, 1978.

Bowles, B. Educational Pressure Groups and the Legislative Process in California, 1945-1966. Unpublished doctoral dissertation at the Claremont Graduate School, Claremont, Calif., 1965.

Brzezinski, E. Testing in Michigan: A Twenty-Year Perspective. Michigan School Testing Service, University of Michigan, 1977.

Buros, O. Tests in Print. Vol. I and Vol. II. Highland Park, N. J.: Gryphon Press, 1972.

Campbell, R., Cunningham, L., & McPhee, R. The Organization and Control of American Education. Columbus, Ohio: E. Merrick Books, 1965.

Carver, R. P. The Coleman Report: Using Inappropriately Designed Achievement Tests. American Educational Research Journal, 1975, 12, 77-86.

Center for the Study of Education. Determining How Well a Test Measures your Objectives. Los Angeles, April, 1975.

Clasby, M. Laws, Tests and Schooling: Changing Contexts for Educational Decision-Making. New York: Syracuse Center for Policy Research, October 1976.

Cohen, D. Reforming School Politics. Harvard Educational Review, 48 (4), 1978.

Dahl, R., & Lindblom, C. Politics, Economics and Welfare. New York: Harper & Brothers, 1953.

Dahl, R. A Preface to Democratic Theory. Chicago: University of Chicago, 1956.

Ebel, R. State Testing Program: Status Problems and Prospects. In Tests, Measurement and Evaluation. ERIC Clearinghouse,Dec. 1974.

Friedman, M. Capitalism and Freedom. Chicago: University of Chicago Press, 1962.

Friendly, H. The Federal Administrative Agencies. Cambridge, Mass.: Harvard University Press, 1962.

Fromm, E. The Sane Society. New York: Reinhart, 1955.

Galbraith, J. The New Industrial State. 2nd ed. Boston: Houghton, 1971.

Gilbert, H. Teacher Selection Policies and Procedures in Large Public School Systems in the Coastal States. Board of Education, New York City, 1966.

Glaser, B., & Strauss, D. The Discovery of Grounded Theory. Chicago: Aldine Publishing Co., 1973.

Haney, W. Standards for Tests and Test Use. National Consortium on Testing, Staff Circular #3, October 1978.

Haney, W., & Madaus, G. Making Sense of the Competency Testing Movement. Harvard Educational Review, 48 (4), 1978.

Hector, L. Problems of the CAB and the Independent Regulatory Commissions. Yale Law Review, 69, 939, May 1960.

Hobbs, N. The Future of Children. San Francisco: Jossey Bass Publications, 1975.

Hoffman, B. The Tyranny of Testing. Crowell-Collier Press, 1962.

Hunter, B. T. Consumer Beware. New York: Simon and Schuster, 1972.

Kariel, H. The Decline of American Pluralism. Stanford, Calif.: Stanford University Press, 1961.

Karier, C. Testing for Order and Control in the Corporate Liberal State. Educational Theory, 22, Spring, 1972.

Kennedy, M., Apling, R., & Neuman, B. The Role of Evaluation and Test Information in Public Schools. Cambridge, Mass.: Huron Institute, 1980.

Kiersh, E. Testing is the Name, Power is the Game. Voice, January 15, 1979.

Kohn, S. The Number Game: How the Testing Industry Operates. In P. Houts (Ed.), The Myth of Measurability. New York: Hart Publishing, 1977.

Kremble, E. Tests Produce as Many New Questions as Answers. American Teacher, October, 1976.

Levine, M. Academic Achievement Tests. American Psychologist, March 1976, 228-238.

Lindblom, C. Politics and Markets. New York: Basic Books, 1977.

Lowi, T. The End of Liberalism. New York: Norton Press, 1964.

Lyon, C., et al. Evaluation and School Districts. Los Angeles: Center for the Study of Evaluation, December, 1978.

Madaus, G. School Effectiveness. New York: McGraw-Hill, 1980.

Madaus, G. Minimum Competency Testing: Unexamined Assumptions and Unexplored Negative Outcomes. New Directions for Testing and Measurement, 2, 1979.

Madaus, G., & Rippey, R. Zeroing in on the STEP Writing Test: What Does it Tell a Teacher? Journal of Educational Measurement, 1966, 3, 19-25.

Madison, J. The Federalist Papers. New York: Mentor Books, 1961.

Marconnit, G. School Legislatures and the School Curriculum. Phi Delta Kappan, January, 1968.

McClelland, D. Testing for Competence rather than Intelligence. American Psychologist, 28, 1, January 1973, 1-14.

McConnel, G. Private Power and American Democracy. New York: Knopf, 1967.

Nairn, A., and Associates. The Reign of E.T.S.: The Corporation that Makes Up Minds. Nader Report. Washington, D.C.: Learning Research Project, 1980.

National Education Association. Research Report on Measurement and Testing. Washington, D.C., 1977.

National School Boards Association. Standardized Achievement Testing. Research Report No. 1, 1977.

Oles, H., & Davis, G. Publishers Violate APA Standards on Test Distribution. Psychological Reports, December 2, 1977, 713-719.

Pipho, C. Minimum Competency Testing in 1978: A Look at State Standards. Phi Delta Kappan, May 1978, 585-588.

Porter, A. Relationships between Testing and the Curriculum. Occasional Paper No. 9, Institute for Research on Teaching, College of Education, Michigan State University, July 1978.

Porter, A., & McDaniels, G. A reassessment of problems in estimating school effects. Paper read at the Annual Meeting of the American Association for the Advancement of Science, March, 1974.

Schrag, J., et al. Education for Handicapped in Rural Areas. Paper presented at the Rural Education Seminar, College Park, Maryland, Mary 1979.

Scherer, F. Industrial Market Structure and Economic Performance. Chicago: Rand McNally, 1970.

Stake, R. The Case Study Method in Social Inquiry. Educational Researcher, Vol. 7, No. 2, Feb. 1978.

Stake, R. Measuring What Learners Learn. Urbana, Ill.: Center for Instructional Research and Curriculum Evaluation. University of Illinois, 1972.

Stodolsky, S. Refining Treatment and Outcome in Early Childhood Education. In H. Walberg & A. Kupans (Eds.), Rethinking Urban Education. San Francisco, Calif.: Jossey Bass, 1972.

Strenio, A. The Testing Trap. New York: Atheneum, 1980.

Thomas, N. Education in National Politics. New York: McKay Co., 1975.

Thurow, L. The Zero-Sum Society. New York: Basic Books, 1980.

Truman, D. The Governmental Process. New York: Knopf, 1951.

Tyler, R. The Use of Tests in Measuring the Effectiveness of Educational Programs, Methods, and Instructional Materials. In R. Tyler & R. M. Wolf (Eds.), Crucial Issues in Testing. Berkeley, Calif.: McCutchan, 1979.

Ward, J. Teachers and Testing. A Survey of Knowledge and Attitudes. A Report of the Research Department of the American Federation of Teachers. AFL-CIO, July 1980.